Genomics of lung cancer

Saber Hosseinabadi, Ali

Chapter 5

Identification of novel fusion genes in lung adenocarcinoma patients

Chapter 5A

Identification and validation of predicted gene fusions: A pilot study to optimize the bio-analytic procedure

Ali Saber[1], Anthonie J. van der Wekken[2], Klaas Kok[3], M. Martijn Terpstra[3], Wim Timens[1], Debora de Jong[1], T. Jeroen. N. Hiltermann [2], Harry J.M. Groen[2], Anke van den Berg[1]

University of Groningen, University Medical Center Groningen, [1]Department of Pathology and Medical Biology, [2]Department of Pulmonary Diseases, [3]Department of Genetics, the Netherlands

Abstract

Introduction: Fusion genes can be formed as the result of chromosomal translocations, deletions or inversions. Recent developments in sequencing techniques and bioinformatics analysis enable researchers to identify and validate novel fusion genes using paired-end RNA sequencing on primary tumor material. The aim of this pilot study is to improve selection of predicted of high confidence fusion genes for follow-up studies

Materials and methods: RNA was isolated from a frozen lung tumor biopsy of a patient with an adenocarcinoma and subjected to paired-end RNA sequencing. Reads were mapped to the reference genome (hg19) and potential fusion genes were called using a fusion prediction tool (deFuse). After applying different filtering steps and manual inspection of the reads using the IGV and the UCSC browser, selected high confidence fusion genes were validated by RT-PCR.

Results: Eighty-five potential fusion genes were called by deFuse. Of these, 18 had a probability of ≥0.95 and 67 had probability <0.95. A further selection of the 18 potentially fusion genes was achieved by excluding predicted fusion genes that were likely to be the result of read-through transcripts of adjacent genes (n=7). Manual inspection of the remaining 11 candidates resulted in a further exclusion of 2 predicted fusion genes, based on lack of split reads and mapping to an intergenic region, respectively. Three of the 9 remaining fusion genes involved the same two genes, i.e. *SCNN1A* and *TNFRSF1A,* but with different breakpoints. Four fusion genes contained a predicted open reading frame (ORF) and were validated by RT-PCR. RT-PCR validation of *SCNN1A* and *TNFRSF1A* using a primer set flanking the three breakpoint regions did yield the expected bands in the tumor sample, but also revealed PCR products in a normal control sample. These predicted fusion genes could have been excluded by an additional filtering step, i.e. removal of fusion genes located in expressed sequence tag (EST) enriched regions. The other three predicted fusion genes did yield the expected bands in the tumor samples and were validated by Sanger sequencing. No PCR products were observed in the normal control sample. This indicated that these three predictions most likely present true gene fusions.

Conclusion: The application and validation of the deFuse fusion gene detection software in a lung cancer specimen is improved by inspection of the resulting fusions for read-through variants, ESTs and spanning reads.

Introduction

Fusion genes are the result of genomic rearrangements that occur due to incorrect DNA repair caused by errors in the DNA repair machinery[1-2]. A number of fusion gene products have been shown to be causally involved in a variety of diseases such as hematological malignancies and childhood sarcomas[2-3]. Fusion genes were detected mainly in leukemia and sarcoma, and appeared to be less common in solid tumors, such as lung cancer. The Philadelphia chromosome was the first identified translocation and is a classical example of a specific gene fusion, i.e. *BCR-ABL1*, in a human neoplasia[4]. It was discovered in chronic myelogenous leukemia (CML) and plays an important role in tumor cell survival through inhibition of apoptosis[4-5].

In B-cell lymphoma, many characteristic chromosomal translocations have been identified, but unlike leukemia and sarcoma, these translocations do not result in fusion genes. Instead these B-cell lymphoma-specific translocations result in overexpression of the target gene due to juxtaposition of the gene to a genomic region that contains strong enhancers, such as the immunoglobulin heavy and light gene enhancers. An example is the chromosomal translocation in Burkitt lymphoma, which places the *MYC* proto-oncogene under control of the immunoglobulin (Ig) heavy or light chain gene enhancers[2, 6]. The *MYC* translocation results in transcriptional activation of the *MYC* oncogene by Ig regulatory elements[2].

In solid tumors, the presence of fusion genes remained undetected until recently. *EML4-ALK* was the first fusion gene to be discovered in non-small cell lung cancer[7] with a prevalence of around 3-6%[8]. Recurrent fusion genes may represent potential treatment targets as the tumor cells are often dependent on signaling cascades activated by the fusion products for growth and survival[9]. This dependency is referred to as "oncogene addiction"[10]. These driver genes are often kinases and they represent novel therapeutic targets for the treatment of lung cancer. Targeted treatment is already available for several of these fusion products. An example is crizotinib for *EML4-ALK* in NSCLC patients[11]. In phase I and II clinical trials[12-13], lung cancer patients harboring *ROS1* and *RET* translocations showed good responses to crizotinib and cabozantinib, respectively.

Developments in sequencing technologies have created high throughput approaches to identify a variety of genomic aberrations in high throughput

approaches[14-16]. Transcriptome sequencing allows quantification of gene expression, identification of splice variants, identification of 5' and 3' ends of transcripts, detection of novel transcripts and discovery of novel fusion products[17-18]. Comprehensive genome and transcriptome sequencing in combination with the development of new algorithms has resulted in the identification of several novel fusion genes[19-20].

We used deFuse, a computational algorithm for discovery of fusion transcripts on paired-end whole transcriptome sequencing data. Information from split reads (reads that cover the fusion boundary) and discordant reads (mate pair reads that map to two distinct genomic regions) are used for prediction of fusion transcripts. DeFuse generates multiple confidence measures to estimate the probability of the predicted fusions and predicts the presence of an ORF in the fusion products[9]. The aim of this chapter is to explore the significance and relevance of the criteria generated by deFuse to select true gene fusions and potentially identify novel fusion genes in a lung cancer sample.

Materials and Methods

Patient

A Fresh frozen tissue sample was obtained from a lung adenocarcinoma patient. A normal lung tissue sample was used as a negative control for validation by RT-PCR. Written informed consent was obtained from the patient.

RNA isolation

Total RNA was isolated using TRIzol following the standard protocol provided by the manufacturer (Life technologies, Carlsbad, USA). The concentration of the RNA samples was measured on a NanoDrop (Thermo Fisher Scientific Inc., Waltham, USA).

Transcriptome sequencing and fusion detection

The truSeq RNA kit (Illumina, San Diego, USA) was used to prepare the library for paired-end RNA sequencing starting with 500ng of RNA. Paired-end reads of 100nt were produced on the Hiseq2500 (Illumina, San Diego, USA). Reads were mapped to the reference genome using deFuse (v.0.6.1) to predict fusion transcripts[9]. DeFuse aligns the reads through an automated process which incorporates SAMtools[21], bowtie[22], BLAT[23] and GMAP[24].

Filtering steps

DeFuse can generate several confidence measures such as number of split and spanning reads, EST island adjusted percent identity score, genome adjusted percent identity score, split-span p-value, a probability score and some other confidence scores. These scores can be used to identify the most likely fusion gene candidates. Besides these criteria included within the deFuse output file, the split read sequence identified by deFuse can be checked manually using a custom track on IGV. Moreover, correct mapping of the sequence of the predicted fusion candidates has been confirmed using the University of California Santa Cruz (UCSC) genome browser.

Based on the initial evaluation of different confidence scores, the optimised filtering steps applied included: probability score ≥0.95; not being derived from read-through products according to deFuse; reliable split and spanning reads using a custom track on IGV[25]; not originating from intergenic regions or read-through transcripts according to the mapping of the fusion gene sequence to the UCSC genome browser.

Validation of the fusion products by RT-PCR

cDNA was synthesized with Superscript II reverse transcriptase and random primers according to the manufacturer's protocol using 500ng of total RNA (Invitrogen, Carlsbad, USA). PCR was performed using 10ng cDNA as input in a final volume of 30µl containing 1x PCR buffer, 0.2µl Tag DNA polymerase (5unit/µl), $MgCl_2$ (final concentration 1.5mM) (Invitrogen, Carlsbad, USA) and 500nM primers designed using Clone Manager Suite (Sci-Ed Software, Morrisville, USA) (Table 1). Amplification was for 35 cycles using a thermocycler (Bio-Rad, Hercules, USA). PCR products were analyzed on a 1% agarose gel. Gel pictures were captured using Gel Doc XR+ System (Bio-Rad, Hercules, USA). PCR products were purified using Zymoclean™ Gel DNA Recovery Kit (Zymo research, Irvine, USA) and sequenced at LGC Genomics (Berlin, Germany). TOPO® TA Cloning® Kit (Invitrogen, Carlsbad, USA) was used to clone the *SLC10A7-TTC29* fusion product according to the manufacturer's protocol. Five independent colonies with the appropriate insert size were sent for Sanger sequencing (LGC Genomics, Berlin, Germany).

Table 1: RT-PCR primer sets for validation of 4 gene fusions with a predicted ORF.

Name	Primer sequence	Annealing (°C)	Expected product size (bp)	Transcript Ensembl ID
RAB35-E1-F	5'-CTTCAAGCTGCTCATCATCG-3'	55	468	ENST00000538903
ATP2A2-E8-R	5'-TGAACCGGGTCATTGAAGTG-3'			ENST00000308664
STAB2-E17-F	5'-CATGGCCAACCAGCTCATAC-3'	55	443	ENST00000388887
NUAK1-E3-R	5'-AGACGATCTGCCGGAAGAAG-3'			ENST00000261402
SLC10A7-E1-F	5'-CTGGTTCATGGTCGGAATAG-3'	55	630/591	ENST00000335472
TTC29-E11-R	5'-GTACTTGCTCTCACCAAATC-3'			ENST00000513335
SCNN1A-E10-F	5'-AGATGCTATCGCGACAGAAC-3'	55	691/624/562/537/495/ 470/441/312/287	ENST00000360168
TNFRSF1A-E4-R	5'-GGTCCACTGTGCAAGAAGAG-3'			ENST00000162749

Results

RNA sequencing and fusion prediction

Paired-end RNA sequencing of pulmonary adenocarcinoma resulted in more than 42 million reads of which over 25 million reads were on target. Eighty-five fusions were predicted in the lung tumor sample by deFuse using the standard settings. Of these, 67 had a probability of <0.95 genes, whereas 18 had a probability of ≥0.95 (Figure 1).

As a quality check we first investigated a number of predicted fusions with a probability of less than 0.95. Manual inspection of these 67 fusion genes using our custom-made track in IGV indicated lack of split and spanning reads for 18 of them. Therefore, these were unlikely to represent true fusion transcripts. A further manual Inspection on UCSC genome browser revealed that 21 of the fusion genes mapped to intergenic regions and 9 fusion products appeared to be a read-through. Based on these results, we decided not to inspect the remaining 19 predicted fusion genes with a probability of <0.95. During this process we noticed that several of the false positive predictions could have been excluded based on the prediction by deFuse as being a read-through transcript. Examples of read-through and intergenic fusion products, as well as fusion products with sufficient and insufficient split reads are shown in Figure 2.

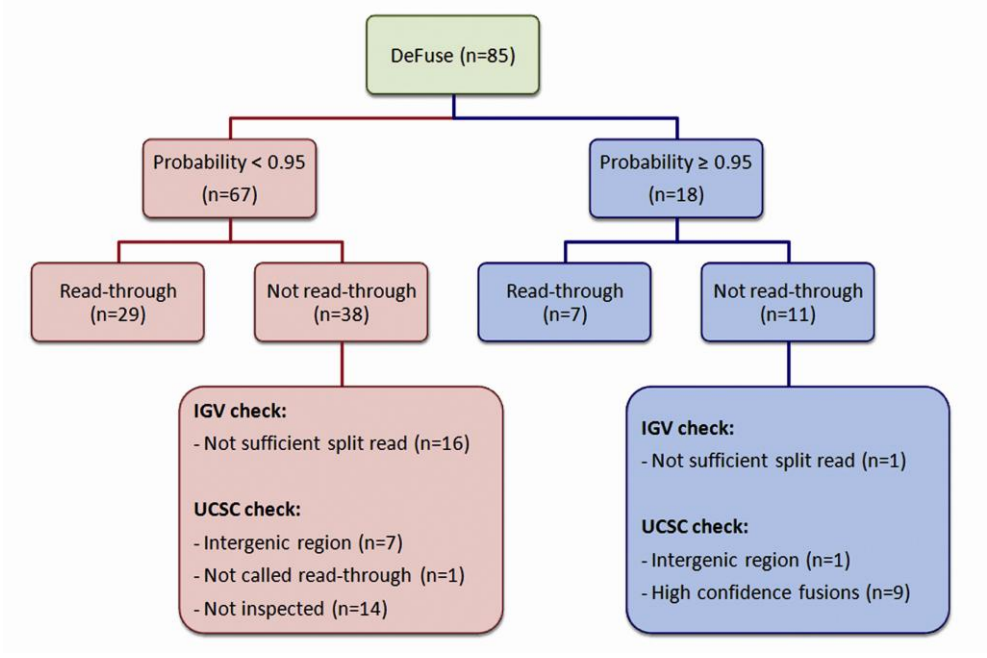

Figure 1: Diagram shows the results of the deFuse filtering and manual inspection steps. Left wing (red) shows results of the initial inspection of the fusions with probability <0.95. Based on these analysis we concluded that the probability score was a good first criterion to exclude false positive predictions. Moreover, we noticed that a second filtering criterion indicated by the deFuse algorithm, i.e. prediction for being a read-through transcript, can correctly remove additional unlikely fusion gene candidates. This second criterion was implemented in the right wing (blue) of the graph to select the most likely candidate fusion genes for RT-PCR validation.

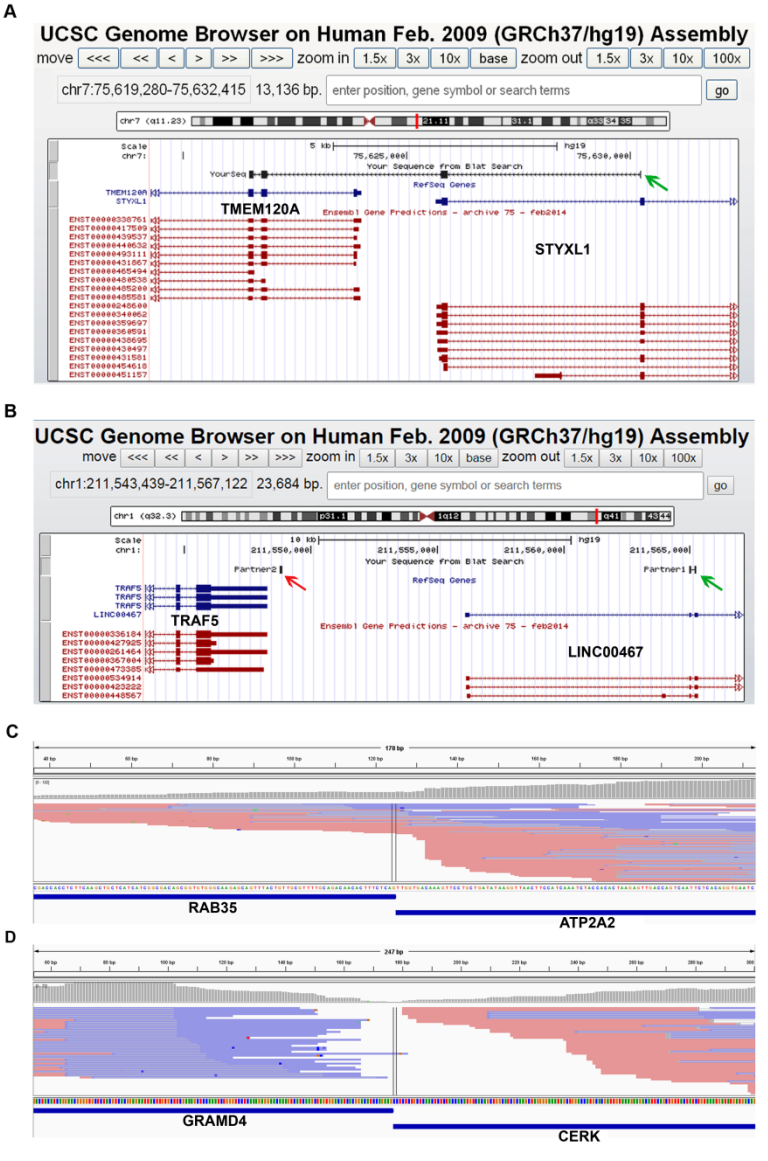

Figure 2: Example of a high confidence fusion and two false positive gene fusions. **A)** Snapshot of *STYXL1-TMEM120A* on the UCSC browser that shows a read-through transcript. Green arrow shows the transcript predicted as fusion by deFuse. **B)** Snapshot of *LINC00267-TRAF5* fusion on the UCSC genome browser showing the sequence (red arrow) which had been annotated as being part of the *TRAF5* gene, but actually maps to an intergenic region in close proximity to this gene (Red arrow). Green arrow shows the sequence that was correctly annotated as the *LINC00267* gene. **C)** Snapshot of IGV of *RAB35-ATP2A2* fusion with sufficient split reads covering fusion boundary. Mixed red and blue area shows paired end reads generated by RNA sequencing. The line in the middle indicates fusion boundary. **D)** Snapshot of IGV for a predicted fusion (*GRAMD4-CERK*) with insufficient split reads covering fusion boundary is an indicative of a false positive gene fusion.

Based on the above described quality check, we used deFuse read-through prediction as a second filter for predicted fusion genes in addition to the probability score of ≥0.95. Seven of the 18 fusion genes were excluded based on this criterion. Inspection of the paired-end reads of the remaining 11 fusion genes using IGV in combination with the custom-made track of the sequence of the predicted breakpoint fusion regions, indicated lack of reliable split and spanning reads for one predicted fusion gene. This fusion was also excluded from further analysis. A further quality check of the remaining fusions using the UCSC genome browser revealed that one predicted fusion mapped to an intergenic region, which was incorrectly annotated as being the more upstream gene by deFuse. This predicted gene fusion was also excluded. Three of the remaining 9 predicted gene fusions involved the same two genes, i.e. *SCNN1A* and *TNFRSF1A,* but with different breakpoints (Figure 3A), for the other six, i.e. *RAB35-ATP2A2*, *STAB2-NUAK1*, *SLC10A7-TTC29*, *KMT2B-AC002115.9, PIAS1-SLC24A6* and *ZNF827-ARHGAP10*, one breakpoint region was observed for each gene. Validation experiments were performed only for the four gene fusions with an ORF as predicted by deFuse.

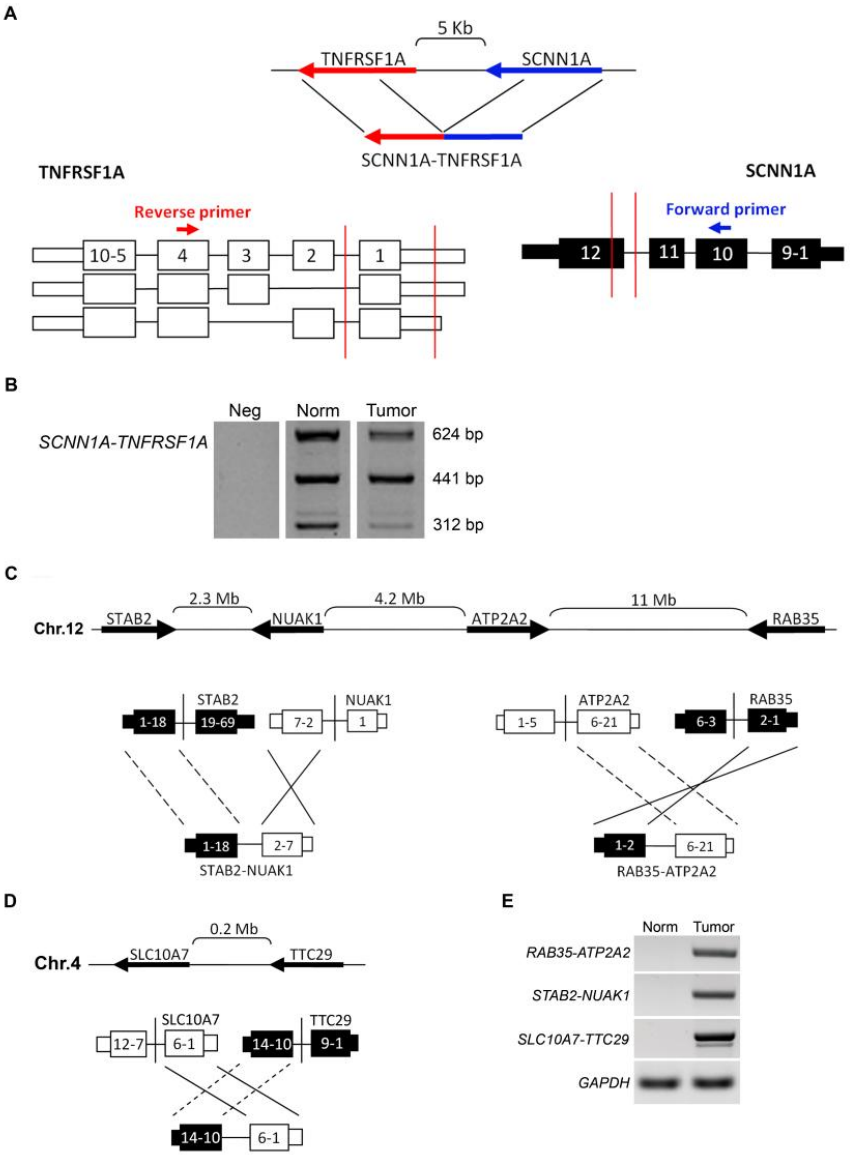

Figure 3: Schematic representation of fusion gene products and validation of four fusion products with a predicted ORF using RT-PCR. **A)** Schematic picture of two genes involved in the predicted *SCNN1A-TNFRSF1A* fusion gene transcript and their orientation in the genome. Red bars show different breakpoints detected by RNA-seq. Position of the primers are shown in the picture. **B)** Validation of *SCNN1A-TNFRSF1A* fusion transcript by RT-PCR. The three fusion products were detected in both the tumor as well as in the normal sample. **C)** Two gene fusions (*RAB35-ATP2A2* and *STAB2-NUAK1*) clustered in an approximately 18Mb region on chromosome 12. Both fusions were result of an inversion. **D)** Schematic picture of *SLC10A7-TTC29* gene fusion which was result of an eversion. **E)** Validation of *RAB35-ATP2A2*, *STAB2-NUAK1* and *SLC10A7-TTC29* fusion product in tumor cDNA by RT-PCR.

Validation of fusion genes with predicted ORF by RT-PCR

Four fusion genes (*SCNN1A-TNFRSF1A*, *RAB35-ATP2A2*, *STAB2-NUAK1* and *SLC10A7-TTC29*) were validated by RT-PCR. Based on different breakpoints detected by RNA-seq and the different splice variants of *TNFRSF1A*, a range of fusion products for the predicted *SCNN1A-TNFRSF1A* fusion gene was expected (Figure 3B). PCR products within the expected range were indeed confirmed in the tumor sample, but a similar pattern was observed in a normal control lung sample, indicative of a false positive fusion prediction. These fusion products were probably derived from read-through transcripts, which had been missed in the initial manual inspection. The fusion product was derived after splicing of the intergenic region between exon 11 of the *SCNN1A* gene and exon 2 of the *TNFRSF1A* gene.

Two fusion products, i.e. *STAB2-NUAK1* and *RAB35-ATP2A2*, were confined to a region of approximately 18 megabases (Mb) on chromosome 12 and involved four genes (Figure 3C). Both fusion products were the result of an inversion. In one of the fusion products exon 18 of the *STAB2* gene was fused to exon 2 of the *NUAK1* gene. In the second fusion product exon 2 of the *RAB35* gene was fused to exon 6 of the *ATP2A2* gene. The fourth fusion, i.e. *SLC10A7-TTC29*, was the result of an eversion in which exon 6 of the *SLC10A7* gene was fused to exon 10 of the *TTC29* gene (Figure 3D). The expected PCR product sizes for all of these fusion genes were identified in the tumor samples and not in the normal sample, suggesting that they are true fusion genes (Figure 3E). Sanger sequencing of the RT-PCR products of these three fusions confirmed the fusion gene sequence as predicted by deFuse.

Discussion

Protein fusion products can act as oncogenic drivers in human cancers[3-4, 7, 26]. DeFuse is a computational algorithm that predicts fusion transcripts from paired-end RNA sequencing data. Here, we aimed to improve selection procedure of predicted fusion genes to allow a more optimal selection of high confidence fusions for further follow-up studies.

A high number of false positive predicted fusion genes involved a gene locus and a second sequence fragment that actually mapped to an intergenic region. These intergenic sequences were annotated to the closest neighboring gene by deFuse resulting in an incorrect calling of fusion genes. This problem is probably

caused by allowing deFuse to call gene fusions in intergenic regions to decrease the chance of missing potential fusions[9]. It seems that this feature can generate a lot of false positive fusion gene predictions and might not be optimal in the search for genuine novel fusion products. Restricting calling of gene fusions to regions within the boundaries of known exons would obviously decrease the number of false positives, but might result in loss of true fusion genes. The probability of being true fusion genes based on the data generated by deFuse appears to be an efficient primary filter to eliminate most of the false positives. In addition, the IGV check using the custom track with the sequences of the breakpoint regions of all predicted fusion genes was a valuable tool to exclude additional falsely predicted gene fusions.

We found several read-through transcripts in our data. Some of these transcripts had been identified already as read-through by the deFuse program, while some others were noted during manual check using UCSC browser or after validation by RT-PCR. These fusion transcripts are the result of a mechanism called "intergenic splicing"[27]. These fusions are also referred to as transcription-induced chimeras or read-through sequences. Read-through transcripts are probably generated when transcription fails to stop at the end of a gene locus and continues until the termination site of the next adjacent gene[28-29]. Read-through sequences usually contain exons of adjacent genes[27]. Overall, excluding read-through transcripts in combination with manual check using IGV on a custom made track and the UCSC browser is an efficient approach to exclude false positive fusions.

In recent years, a number of computational methods have been developed for the detection of fusion transcripts[9, 30-34]. These methods show high sensitivity for detecting artifacts that can be induced by experimental procedures during library preparation and amplification steps prior to RNA sequencing[27]. Thus, selection of alignment software for the detection of fusions and being aware of the type of filters they utilize to remove false positive predictions is important. In addition, close distance (for example <50Kb) between two adjacent genes might be a good sign to consider a transcript as a potential read-through. All together, increasing the specificity can reduce the sensitivity of the method and vice versa. Therefore, we are always on the edge of adding a false positive or missing a true fusion transcript in this type of analysis. Manual check for the selected fusions for validation remains crucial to exclude false positives.

The presence of *SCNN1A-TNFRSF1A* fusion PCR products in a normal sample indicated that this prediction was not valid. So, both deFuse and manual inspection using UCSC genome browser did not identify this predicted fusion gene as being the result of a read-though transcript. This indicates that an additional filtering step should be applied. Therefore, we added an additional filter that could remove predicted fusions that map to EST regions with a probability ≥0.95. Applying this filter resulted in elimination of the *SCNN1A-TNFRSF1A* fusion, without affecting calling of true fusion gene predictions. Following this strategy we were able to reliably distinguish true gene fusions from false positive ones (Figure 4).

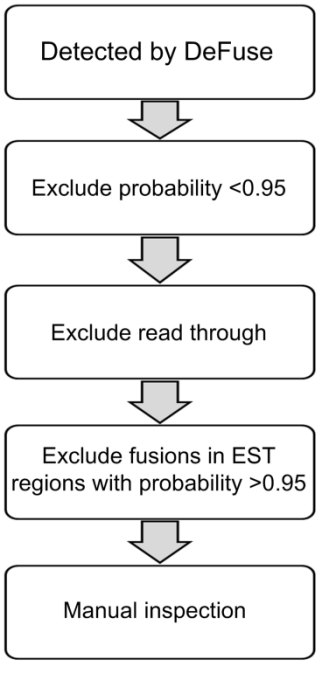

Figure 4: Proposed filtering steps to remove false positive gene fusion predictions by deFuse.

In conclusion, we set up a step-wise approach to select fusion genes with high confidence based on an initial calling made by deFuse and adding additional custom selection steps using the IGV in combination with a custom made track and the UCSC genome browser.

References

1. Aplan PD. Causes of oncogenic chromosomal translocation. *Trends Genet* 2006;22:46-55.
2. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007;7:233-245.
3. Peters TL, Kumar V, Polikepahad S, et al. BCOR-CCNB3 fusions are frequent in undifferentiated sarcomas of male children. *Mod Pathol* 2015;28:575-586.
4. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* 1973;243:290-293.
5. Fernandez-Luna JL. Bcr-Abl and inhibition of apoptosis in chronic myelogenous leukemia cells. *Apoptosis* 2000;5:315-318.
6. Parker BC, Zhang W. Fusion genes in solid tumors: an emerging target for cancer diagnosis and treatment. *Chin J Cancer* 2013;32:594-603.
7. Soda M, Choi YL, Enomoto M, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 2007;448:561-566.
8. Thunnissen E, Bubendorf L, Dietel M, et al. EML4-ALK testing in non-small cell carcinomas of the lung: a review with recommendations. *Virchows Arch* 2012;461:245-257.
9. McPherson A, Hormozdiari F, Zayed A, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* 2011;7:e1001138.
10. Torti D, Trusolino L. Oncogene addiction as a foundational rationale for targeted anti-cancer therapy: promises and perils. *EMBO Mol Med* 2011;3:623-636.
11. Cui JJ, Tran-Dube M, Shen H, et al. Structure based drug design of crizotinib (PF-02341066), a potent and selective dual inhibitor of mesenchymal-epithelial transition factor (c-MET) kinase and anaplastic lymphoma kinase (ALK). *J Med Chem* 2011;54:6342-6363.
12. Shaw AT, Ou SH, Bang YJ, et al. Crizotinib in ROS1-rearranged non-small-cell lung cancer. *N Engl J Med* 2014;371:1963-1971.
13. Drilon A, Wang L, Hasanovic A, et al. Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer Discov* 2013;3:630-635.
14. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012;150:1107-1120.
15. Zhang J, Fujimoto J, Wedge DC, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* 2014;346:256-259.
16. TCGA. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 2014;511:543-550.
17. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;10:57-63.
18. Nagalakshmi U, Waern K, Snyder M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr Protoc Mol Biol* 2010;Chapter 4:Unit 4 11 11-13.
19. Kohno T, Ichikawa H, Totoki Y, et al. KIF5B-RET fusions in lung adenocarcinoma. *Nat Med* 2012;18:375-377.
20. Ju YS, Lee WC, Shin JY, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res* 2012;22:436-445.
21. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078-2079.
22. Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 2009;10:R25.
23. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome Res* 2002;12:656-664.
24. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 2005;21:1859-1875.
25. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 2013;14:178-192.

26. Tognon C, Knezevich SR, Huntsman D, et al. Expression of the ETV6-NTRK3 gene fusion as a primary event in human secretory breast carcinoma. *Cancer Cell* 2002;2:367-376.
27. Llorens-Rico V, Serrano L, Lluch-Senar M. Assessing the hodgepodge of non-mapped reads in bacterial transcriptomes: real or artifactual RNA chimeras? *BMC Genomics* 2014;15:633.
28. Parra G, Reymond A, Dabbouseh N, et al. Tandem chimerism as a means to increase protein complexity in the human genome. *Genome Res* 2006;16:37-44.
29. Akiva P, Toporik A, Edelheit S, et al. Transcription-mediated gene fusion in the human genome. *Genome Res* 2006;16:30-36.
30. Li Y, Chien J, Smith DI, et al. FusionHunter: identifying fusion transcripts in cancer using paired-end RNA-seq. *Bioinformatics* 2011;27:1708-1710.
31. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* 2011;27:2903-2904.
32. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. *Genome Biol* 2011;12:R72.
33. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15-21.
34. Liu C, Ma J, Chang CJ, et al. FusionQ: a novel approach for gene fusion detection and quantification from paired-end RNA-Seq. *BMC Bioinformatics* 2013;14:193.